We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

BOOK CITATION INDEX
CLARIVATE ANALYTICS
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Traceability Technology for Sudden Water Pollution Accidents in Rivers

Weihong Liao, Zhiguo Gan, Jiabiao Wang and Hezhen Zheng

Abstract

The traceability technology for sudden water pollution accidents can be used for fast, accurate identification of a pollution source in the river. A correlation optimization model with the pollution source position and release time as its parameters is established based on hydrodynamic calculation and on the coupling relationship between forward concentration probability density and backward position probability density; and the solution of the model is realized by using a differential evolution algorithm (DEA). A coupled probability density method is to convert the traceability of a sudden water pollution accident into the optimization of two minimum values. This method is simple in principle and easy in solution, realizing the decoupling of parameter of the pollution source. The concept of gradient is introduced to the differential evolution algorithm, improving the efficiency of searching process. The proposed method of traceability was applied to the emergency demonstration project of the SNWDMRP. The results indicate that the model has good efficiency of traceability and high simulation precision and that traceability results have a certain guiding significance to the emergent regulation and control of sudden water pollution events in a river.

Keywords: coupled probability density function, correlation optimization model, parameter decoupling, traceability, emergency, river

1. Introduction

Frequent occurrence of various kinds of water pollution events makes the research of pollutant traceability receive increasing attention, especially for long-distance water diversion projects including the SNWDMRP. The uncertainty of sudden water pollution events increases the difficulty in coping with pollution urgently. Finding out the source of water pollution is a prerequisite for realizing water quality prediction and water body pollution control. Therefore, the first task of urgently coping with sudden water pollution is to determine the source of pollution at the first time after such event, make a reasonable emergency handling plan on the basis of pollution source intensity and the occurrence place and time which have been determined, and meanwhile, provide preconditions for sudden water pollution prediction and early warning [1, 2]. The aforesaid traceability technology infers the position and time of occurrence as well as pollution source

intensity through research on the transfer and conversion laws of pollutants in a river and on the basis of a monitored pollutant concentration process, realizing the reconstitution of the pollution event and playing an important role in the emergent regulation and control process of sudden water pollution events [3].

2. Basic traceability principle of sudden water pollution

Essentially, both pollutant concentration prediction and pollutant traceability fall into the process of pollutant traceability. To be more precise, the prediction of pollutant concentration distribution can be taken as the forward track process of pollutants, while the traceability of pollutants the process of backward track and traceability, as shown in **Figure 1**.

According to the information on a pollution event, the process of determining the concentration of the pollutant during the event by using the simulation technology in Chapter 2 is forward track of pollutant transfer and whereabouts, while the tack of the related information on the occurrence process and source of the event by using observed data including pollutant concentration distribution is the reversal of the forward track. Compared with the prediction of pollutant concentration distribution, the traceability of pollutant is obviously more complex, not only involving the process of track and backward reconstitution of the event but also having nonlinearity and ill-posedness as the reverse problem of prediction.

According to the connotation of sudden water pollution traceability, the definition of such traceability can be divided into five categories: The first category of traceability refers to reconstitution of unknown coefficients in a pollutant transport model according to some (observed) information of temporal and spatial distribution of a known pollutant, including longitudinal dispersion coefficient, lateral dispersion coefficient, and degradation coefficient, and the research is called a problem on parameter identification research; the second category refers to inference of the right pollution source (collection) item in the model by using observed values, including pollution source position and release intensity and time, and such traceability is also called an identification problem of the pollution source (collection) item (called a traceability problem for short); the third category is reverse inference of initial conditions on the basis of known information, and such traceability is also called a backward-time research problem; the fourth category means to infer the type or parameters of boundary conditions according to known information, namely, the reverse inference research of boundary conditions; and the fifth category is the combination of the four categories above. The
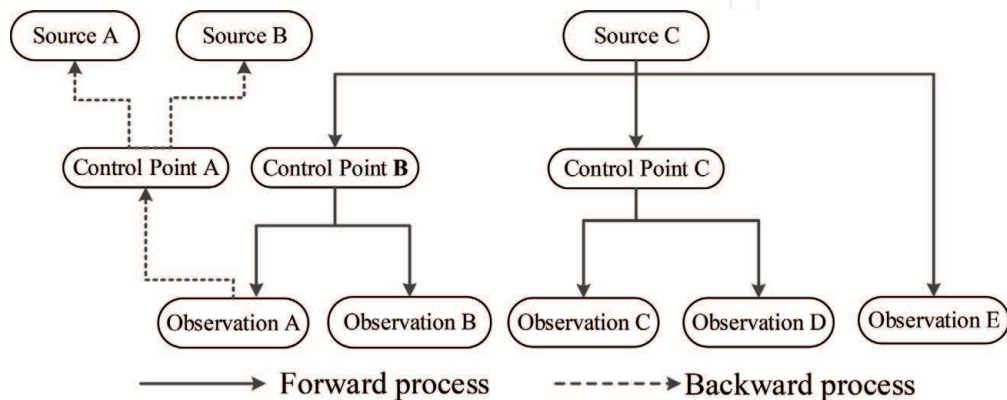

Figure 1.
Traceability process of sudden water pollution.

pollutant traceability technology studied in this chapter is mainly about the second category traceability problem, that is, the identification problem of source (collection) items.

Attention has being paid gradually to the research of the traceability of sudden water pollution in recent years. There are two kinds of common research methods: deterministic method and probabilistic method [4, 5]. The deterministic method mainly includes a regularization method [6, 7], a trial and error method [8], a least square method [9, 10], and other optimization methods. These methods have a clear physical meaning. A single solution can be obtained by using them, and the error is often great if information is not accurate. However, the probabilistic method is a kind of random method based on Bayes inference and Markov chain Monte Carlo (MCMC) sampling, providing multiple reliable alternative results depending on the collection of distribution information on random variables [4, 11, 12]. On the basis of common traceability methods, this chapter introduces a new fast traceability method for sudden pollution in a channel, which combines the deterministic and probabilistic methods [1] and provides multiple reliable alternative results for urgently coping with sudden pollution without depending on distribution information on random variables.

3. Common traceability methods

3.1 The deterministic method

The deterministic method is developed mainly by reference to the traceability research of underground water. Perform research on pollutant concentration distribution in a simulation event by using a pollutant transfer and diffusion model, set up an optimization model which takes the error square sum between the simulation result and the result actually measured as an objective function, solve the objective function of the optimization model by using a deterministic algorithm, and then seek for, via iteration, the calculation result which best matches the values actually observed. Typically, the deterministic method includes the regularization method, trial and error method, correlation regression analysis, and the like, as well as some heuristic algorithms including genetic algorithm and simulated annealing algorithm. Regarding the tedious problems of the trial and error method, Han et al. [8] performed the inversion by using "local and basic expansion of a forward problem" and the optimal control solution by converting traceability into a minimum value problem. Jin and Chen [13] obtained a response relation between the concentration of pollutants in a channel via variation of an objective function established by using the Laplace transform. On such basis, they determine the pollution source intensity of an upstream cross section under environmental capacity control via reasoning. Chen et al. [14] transformed the calculation expression of instantaneous discharge concentration at a 1D single-point source, obtaining a linear regression model. And via regression analysis, they determined the release position, time, and intensity of the pollution source as well as the longitudinal dispersion coefficient of a channel. Min et al. [15] carried out, by using genetic algorithm, the research, respectively, on the identification problem of multiple parameters including the flow velocity, diffusion coefficient, and attenuation coefficient of a 1D river, as well as on the identification problem of right items of a 1D convection-diffusion equation. This part presents a simple introduction of the deterministic method by taking the correlation regression analysis method as an example.

When pollutants are released instantaneously, the analytic solution of the 1D convection-diffusion equation can be further determined via Fourier transform or dimensional analysis. Assuming that the occurrence position and time of pollution are x_0 and t_0, respectively, then we obtain

$$C = \frac{M}{\sqrt{4\pi D(t - t_0)}} \exp\left(-\frac{((x - x_0) - u(t - t_0))^2}{4D(t - t_0)}\right) \tag{1}$$

where M is the average intensity of the initial surface source of the released pollutant along a cross-section, g/m^2.

After taking logarithmic transformation of both ends of Equation (1), we obtain

$$\ln C = aX + b \tag{2}$$

where $X = ((x - x_0) - u(t - t_0))^2$, $a = -\frac{1}{4D(t-t_0)}$, and $b = \ln \frac{M}{\sqrt{4\pi D(t-t_0)}}$.

For a fixed pollution event, the occurrence position x_0 and the time t_0 are constants. Therefore, if the concentration of multiple cross sections can be obtained at the same time and the control time parameter $(t - t_0)$ is a constant, then it is easy to judge that $\ln C$ and X (i.e., the observation position x) meet a linear relation; and a linearly dependent optimization model can be built by using a regression analysis method [14]:

$$R = \frac{\sum_{i=1}^{n}\left(\ln C_i - \overline{\ln C}\right)\left(X_i - \overline{X}\right)}{\sqrt{\sum_{i=1}^{n}\left(\ln C_i - \overline{\ln C}\right)^2}\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}} \tag{3}$$

With the position of the monitored cross section as a variable, after selection of x_0 and t_0, the concentration measured C_i at different monitored cross sections x_i at the same time and corresponding to X_i, and at the monitored position corresponding to the same time are obtained via calculation. Theoretically, when x_0 and t_0 are selected as the true position and time of occurrence, R obtained via calculation should be 1.0.

From this, the problem of pollutant traceability is converted into determination of proper x_0 and t_0 via Formula (3), and the R meeting calculation is maximum, that is, 1.

The solution can be completed in five steps:

1. Firstly, calculate a proper $x' = x_0 + u(t - t_0)$, and it is selected to be 1.0 if meeting R. The calculation can be realized via derivation.

2. Calculate X_i corresponding to different observed cross sections at the same time according to known x', fit $\ln C$ and X in combination of Formula (2), calculate the corresponding slope a and intercept b, and therefrom, obtain t_0 via calculation.

3. Calculate the time parameter $t_0 = t + \frac{1}{4Da}$ according to calculated slope a, then $x_0 = x' - u(t - t_0)$ in combination of known x' and finally the intensity of the initial surface source $M = \sqrt{4\pi D(t - t_0)} \exp(b)$ in combination of known b.

4. The intensity of discharged pollutants calculated by using flow cross-section area A at time t_0 at estimated pollution position x_0 is MA.

5. Considering the existence of fitting and observation errors, calculate R via back substitution of obtained x_0 and t_0, then judge whether R is 1. If R is 1 (or the error is very small), then complete traceability calculation. Otherwise go back to Step (2), properly adjust the slope a and intercept b, and then recalculate x_0 and t_0.

From the aforesaid solution process, we can see that the observed data of multiple cross sections are needed at the same time for the model to realize the traceability of the sudden pollution due to a point source.

3.2 The probabilistic method

Because a traceability problem has ill-posedness, when the deterministic method is used for solution, the error in observation or model calculation might result in great deviation in results and distortion in the result of traceability. For this, random devices are introduced to the study of traceability. Among such devices, the one mostly used is the probabilistic method based on Bayes inference and MCMC sampling. Bayes inference is a kind of method based on the theoretical foundation of probability theory, which can reflect the uncertainty of sudden water pollution events in a channel. It solves the posterior probability distribution of such parameters on the basis of making full use of a likelihood function and the priori information of parameters to be determined and then obtains the estimated values of all parameters of a pollution source by respective sampling. This method can offer a random distribution function of the traceability result of a water pollution event. Therefore, the methods based on Bayes inference are mainly used to estimate happening probability of a sudden water pollution event. By using them, we can obtain the posterior probability distribution of the traceability result instead of a single solution, quantitate the uncertainty of the result, and gain more traceability information on the event. In order to acquire the estimated value of the result, Bayes inference should be combined with related sampling methods, such as Markov chain Monte Carlo (MCMC), randomized quasi-Monte Carlo (Monte Carlo, MC), and other sampling methods [11, 12]. Among them, the MC method is an estimation method in which an initial value easily converges to a suboptimal solution no matter whether the initial value deviates from its true value. Therefore, the accuracy rate of the traceability result obtained via this method is not high. The defect in the MC method can be compensated by iterating the mode of Bayes inference's combination with the MC or MCMC method into the obtained distribution function of the traceability result. The MCMC method is a sufficiently long Markov chain acquired via random walk, only by which can the proximity of the sampling result to the posterior distribution of the traceability result be ensured, that is, the limit distribution of Markov chain is used to express the posterior probability density function of the traceability result. For this, the MCMC method expands the application of Bayes inference in the traceability research of environmental pollution events.

Taking Bayesian MCMC method, for example, this part provides a simple introduction of the traceability method in probability statistics. The method takes all variables in a traceability research model as random ones and considers that the solution to the traceability problem of a sudden water pollution event is a probability distribution. Firstly, it converts the priori information of the solution into a priori probability distribution by using Bayes method, then combines with observed data, and finally obtains the posterior probability

distribution of the solution to be determined from the random sampling process of Markov chain by using the likelihood function between observed values and the values calculated via simulation. The probabilistic method includes the following main processes:

1. Model a basic traceability problem, and quantitate the priori information quantity of related traceability solutions (the intensity, occurrence position, and time of a pollution source) by adopting a reasonable probability distribution function.

2. Select and establish a reasonable likelihood function based on a channel water quality coupling simulation model and in combination of the information related to the site of a sudden water pollution event and to the hydrologic data of water space of the event.

3. Obtain a posterior probability distribution of traceability solutions based on the prior probability distribution and the likelihood function.

4. Perform sampling of the posterior probability distribution and acquire the estimated values of traceability solutions to the event.

Firstly, express the probability distribution table of the priori information on the known parameter vector θ before collecting observed data into $p(\theta)$. After obtaining observed data, the posterior distribution of the known parameter acquired via Bayes inference is $p(\theta|d)$, meeting

$$p(\theta|d) = p(\theta)p(d|\theta)/p(d) \tag{4}$$

where θ stands for three parameters of the pollution source; d is the measured value of pollutant concentration.

$p(d)$ is the probability distribution of the measured value. $p(d)$ obviously has no relation with parameter θ. In the case with a known concentration actually measured, $p(d)$ can be understood to be 1. From this, we obtain

$$p(\theta|d) = p(\theta)p(d|\theta) \tag{5}$$

The prior distribution $p(\theta)$ of parameter θ in Formula (5) can be deemed a uniform distribution of a respective parameter within a prior range of values.

$$p(\theta) = \prod U(\theta_i) \tag{6}$$

Therefore, to determine the posterior distribution $p(\theta|d)$ of pollution source parameters, we firstly need to determine distribution $p(d|\theta)$, which can be defined as the likelihood function of a measured and a predicted value. Let d_i, $C_i(x,t|\theta)$, and $p(d_i|\theta)$ to be the measured value, predicted value, and likelihood function of no. i measuring point, respectively, and $\varepsilon_i = d_i - C_i(x,t|\theta)$ to be an error of measurement, $i = 1, 2, ..., N$. Assuming that obedience mean of error ε_i is 0, standard deviation is the normal distribution of ε_i, and all measuring points are independent from each other, then

$$p(d|\theta) = \prod_{i=1}^{N} p(d_i|\theta) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \exp\left(-\sum_{i=1}^{N} \frac{(d_i - C_i(x,t|\theta))^2}{2\sigma^2}\right) \tag{7}$$

Thereby, we obtain the posterior distribution of pollution source parameters $p(\theta|d)$:

$$p(\theta|d) = p(\theta)p(d|\theta) = \prod U(\theta_i) \cdot \frac{1}{\left(\sqrt{2\pi}\sigma\right)^N} \exp\left(-\sum_{i=1}^{N} \frac{(d_i - C_i(x,t|\theta))^2}{2\sigma^2}\right) \quad (8)$$

However, complicated $C_i(x,t|\theta)$ or great model parameter space and number of dimensions make $p(\theta|d)$ very abstract and difficult to be expressed visually. Therefore, direct use of Bayes method almost cannot solve an actual problem directly, which, however, can be solved by means of the MCMC method. According to different transition probability matrixes making up the Markov chain, the MCMC method mainly has the following sampling algorithms: Gibbs sampling algorithm, Metropolis-Hastings algorithm, and self-adaptive Metropolis algorithm. The self-adaptive Metropolis algorithm can converge to a target distribution for any prior distribution of θ, so this algorithm is selected for sampling.

According to the aforesaid derivation process, the main steps for traceability of a sudden pollution event based on Bayes inference and the MCMC method are as follows:

i. Suppose $i = 0$, and then initialize different variables.

ii. Produce and accept random variables, and make up a Markov chain.

    1. The priori parameter θ producing a uniform distribution equals $\theta(m, x_0, t_0)$.

    2. From the produced pollution source parameters, calculate the concentration value at the observing point by using the methods in Chapter 3.

    3. Calculate the likelihood function $p(\theta|d)$ by using Formula (8).

    4. Calculate the acceptance probability of the Markov chain by using the following formula:

$$\alpha = \min\left\{1, \frac{p(\theta^*|d)}{p(\theta_i|d)}\right\}$$

    5. A random number R is produced which uniformly distributes within 0–1. If $R < \alpha$, then accept the parameters for this test and make $\theta_{i+1} = \theta^*$. Otherwise let original parameter $\theta_{i+1} = \theta_i$.

iii. Repeat Steps (1)–(5) until reaching the number of predefined iteration or acquiring the posterior sample number preset for pollution source parameters, and then perform statistics of the posterior distribution law of all parameters and complete the traceability calculation.

From the aforesaid process, we can see that the probability statistics method depends on the priori range of values of parameters and the information of calculation error distribution and that the convergence rate of producing parameter values

randomly by using uniform distribution is very slow if the priori range of values is large. Moreover, the probabilistic method reaches solution based on Markov chain Monte Carlo random sampling, without directivity for making no optimal judgment of existing results, also resulting in a lower efficiency of traceability.

4. Coupled probability density method

4.1 Basic principles and method

According to the introduction and analysis of the common methods mentioned above, the deterministic method has a clear physical meaning but a single solution, resulting in a great error in the case with inaccurate information. Although the probability statistics method can offer multiple reliable alternative results, it depends on the acquisition of distribution information on random variables. For the point of view of research trend, a new generation of traceability method which integrates the deterministic method and the probability method will become the future trend of pollutant traceability research. The coupled probability density method proposed in this chapter combines the deterministic method with the probability method and sets up a sudden water pollution traceability model based on a coupled probability density function (C-PDF). Based on hydrodynamic calculation and considering the observation error of a system, an optimization model is built by taking pollution source position and release time as parameters and through correlation analysis of the forward concentration distribution probability density and backward position probability density of pollutants, and then the solution of such model is realized by using DEA. On such basis, a minimum value optimization model is established based on the forward concentration distribution probability density function of pollutants to determine the intensity of the pollution source.

In case of a sudden event, the majority of pollutants enters a channel in the form of a point source; move and convert with water flowing in the channel. Based on traceability research at home and abroad and the transfer and conversion law of pollutants in a channel, this paper sets up a 1D traceability model for sudden river water pollution depending on a coupled probability density function (C-PDF) and infers the pollution source in the light of observed concentration of pollutants, thus realizing the reconstitution of pollution event. Based on 1D hydrodynamic calculation and considering the observation error of a system, an optimization model is built by taking pollution source position and release time as parameters and through correlation analysis of the forward concentration distribution probability density and backward position probability density of pollutants, and then the solution of such model is realized by using DEA. Meanwhile, a minimum value optimization model is established based on the forward concentration distribution probability density function of pollutants to determine the intensity of the pollution source.

The magnitude of pollutant concentrations at different cross sections in a channel can be expressed as the statistic at such cross sections where tiny substance particles appear at a respective time. Statistically, the high or low probability that substance particles appear at some position can be expressed by a probability function, and the concentration distribution of pollutants in a channel can be also described by some probability density function. On the contrary, without knowing the pollution source, the pollutants observed at some cross section in a river might come from any upstream place, whose probability also can be described by a probability density function. Considering that pollutant traceability is the reverse

problem of pollutant concentration distribution, a probability density function used to describe such distribution is called a forward concentration probability density function. However, a probability density function used to describe the probability that the pollution source is at different positions is called a backward position probability density function. Depending on the definition of the probability density function, a forward probability density function can be obtained after normalization of concentration.

According to $\int_x C(x,t)dx = m_0$, the normalization of $C(x,t)$ can be done, as shown below:

$$c(x,t) = C(x,t)/m_0 \qquad (9)$$

where $c(x,t)$ is the forward concentration probability density value corresponding to $C(x,t)$, with a dimension of m^{-1}, indicating the probability that pollutants appear at cross section x at time t.

According to Neupauer and Wilson's derivations, forward concentration transport and backward position traceability are adjoint processes to each other [16, 17]. For a 1D channel, $P(x_s,t')$ is used to express the probability, judged from an observed cross-section x_d, that a pollution source is at x_s at time t' (i.e., the probability that pollutants are transported from cross-section x_s to cross-section $t_d - t'$ within time x_d, then $P(x_s,t')$ meets the adjoint state equation of a convective diffusion Eq. (10)) and normalization conditions ($P(x_s,t')$ also has a dimension of m^{-1}), as shown in the following formula:

$$-\frac{\partial P(x_s,t')}{\partial t} + \frac{\partial(uP(x_s,t'))}{\partial x} + D\frac{\partial^2 P(x_s,t')}{\partial^2 x} = 0 \qquad (10)$$

$$P(x_d,t_d) = 1 \qquad (11)$$

where t' is the time point for inverse calculation and t_d is the time point for observing pollutant concentration. Formula (11) indicates that the pollution source only can be at the observed cross section when pollutants are not transported ($t' = t_d$) but appear at such cross section.

Like concentration transport process, Formula (11) can be considered a position probability transport process. Similarly, when u is constant, an analytic solution can be obtained, as shown below:

$$P(x_s,t') = \frac{1}{\sqrt{4\pi D(t_d - t')}} \exp\left(-\frac{(x_d\text{-}x_s - u(t_d - t'))^2}{4D(t_d - t')}\right) \qquad (12)$$

We can see that $P(x_s,t')$ and $C(x,t)$ have the same form. Actually, the relation between Formulas (1) and (12) decides that notwithstanding flow field, the relation between $P(x_s,t')$ and $C(x,t)$ can be determined by **Figure 2**, on which the arrows show the direction of transport (advection item).

Pollutant concentration distribution and traceability are inverse problems to each other, but both of the problems follow the same physical law basically. **Figure 2** shows that when $t - t_0 = t_d - t'$, the probability $c(x_d,t)$ that pollutants move from source x_0 to cross-section x_d after time $t - t_0$ equals to the probability, judged by an observer at cross-section x_d, that pollutants move from cross-section x_0 to cross-section x_d after time $t_d - t'$. Namely, if $t - t_0 = t_d - t'$, the following Formula (13) holds
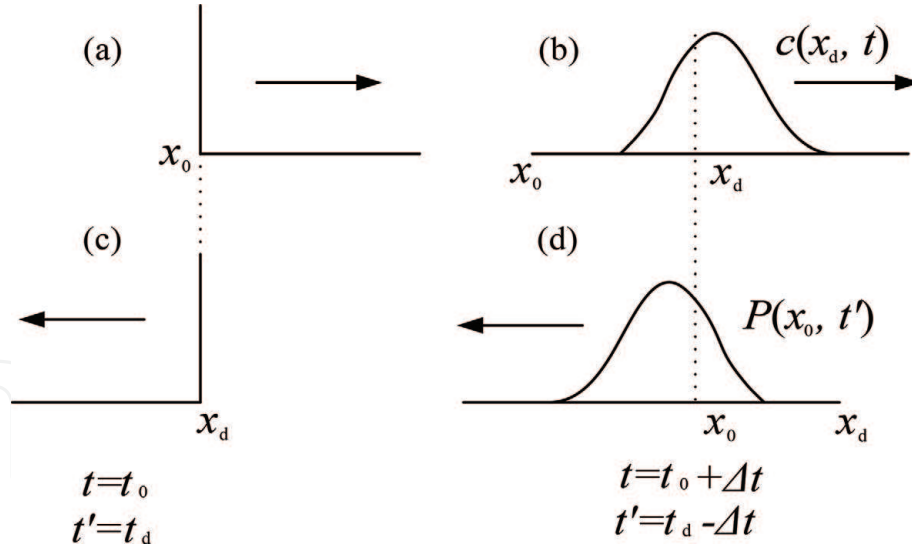
$$P(x_0,t') = c(x_d,t) \qquad (13)$$

Figure 2.
Processes of forward concentration transport and backward position probability transport. a and b are the concentration probability distributions along the river at release time t_o and after Δt, respectively; c and d are the position probability density distributions at monitoring time t_d and of backward traceability for Δt, respectively.

From **Figure 2** and Formula (13), we can see that forward concentration transport process is highly coupled with backward position probability transport process. The two processes are exactly the same except for backward-time calculation direction. Because $P(x_s, t')$ has no relation with the intensity of the pollution source, $P(x_s, t')$ can be directly calculated by using Formula (10) or (12) with $x_s = x_0$ and $t' = t_d + t_0 - t$ given. Therefore, a linearly dependent model can be established by substituting $P(x_s, t')$ for $C(x_d, t)$ based on such coupling relation in Formula (13) to realize traceability calculation.

Assuming that observed concentration series is C_i and that the corresponding calculated position probability density series is P_i, $i = 1, 2, ...n$, according to the aforesaid derivation, we can obtain the expression of the correlation coefficient r of the two series is as follows:

$$r = \frac{\sum_{i=1}^{n}\left(C_i - \overline{C}\right)\left(P_i - \overline{P}\right)}{\sqrt{\sum_{i=1}^{n}\left(C_i - \overline{C}\right)^2}\sqrt{\sum_{i=1}^{n}\left(P_i - \overline{P}\right)^2}} \tag{14}$$

where $\overline{C}$ and $\overline{P}$ are the arithmetic averages of C_i and P_i, respectively.
According to Formulas (13) and (14), we can set up the following function:

$$\min(abs(1 - r)) \tag{15}$$

The constraint condition is the value range of x_0 and t_0, which is given by priori information and generally is obtained via field survey or existing information. See Formulas (16) and (17):

$$x_{0\min} \leq x_0 \leq x_{0\max} \tag{16}$$

$$t_{0\min} \leq t_0 \leq t_{0\max} \tag{17}$$

Through solution of the aforesaid optimization model, we can obtain the release position x_0 and time t_0 of the pollution source.

The optimization model realizes the decoupling of pollution source intensity and time parameter, so determine position and time parameters firstly and then the intensity of the nonpoint pollution source m_0. Considering that the occurrence position and time of the pollution source have been determined and that the range of pollution source intensity m_0 can be roughly determined by the previous model, the forward probability density function can be used to build an optimization model to calculate pollution source intensity. See Formulas (18) and (19).

$$\text{Target function } \min\left(\sum \omega_i (m_0 * c_i - C_i)^2\right) \quad (18)$$

$$\text{Constraint condition } m_{0\min} \leq m_0 \leq m_{0\max} \quad (19)$$

where coefficient $\omega_i = \frac{1}{(C_i + 1.0)^2}$ is used to remove small concentration error loss arising from a large concentration difference.

After verification and analysis, the optimization models (18) and (19) can not only solve the intensity of pollution source m_0 rapidly but also have an advantage in observed error filtration.

4.2 Model solution

The model uses a smart differential evolution algorithm (DEA) method for solution, which is similar to the genetic algorithm [1, 18]. Solving a traceability optimization model by using the DEA includes population initialization, mutation, recombination, and selection. Take the first optimization model made up by Formulas (14) and (16), for example. The DEA defines x_0' and t_0' as property elements and produces basic evolutionary individuals, with $F = abs(1 - r)$ as a fitness target function. The detailed steps of solution are as follows:

(1) Population initialization: let the scale of population equal to NP, and then produce the first individual $X_i^{Gen}(x_0, t_0)$ from Formula (20)
($i = 1, 2, ..., Np$, Gen are the individual and evolution generation numbers, respectively):

$$X_i^{Gen} = X_{\min} + rand(0,1)(X_{\max} - X_{\min}) \quad (20)$$

where $rand(0,1)$ stands for a random number within [0,1]; $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of $X_{i\,Gen}$, respectively. After initialization, calculate their fitness values $F(X_i^{Gen}) = abs(1 - r(X_i^{Gen}))$, respectively.

(2) Mutation: select three individuals X_{r1}^{Gen}, X_{r2}^{Gen}, and X_{r3}^{Gen} from population $Gen(1 < Gen < \max Gen)$ via uniform sampling, and then produce a mutation individual V_i^{Gen} from Formula (21):

$$V_i^{Gen} = X_{r1}^{Gen} + CF(X_{r3}^{Gen} - X_{r2}^{Gen}) \quad (21)$$

where CF is a scaling factor, generally $CF \in [0.5, 1]$, $r1, r2, r3 = 1, 2, ..., NP$, and $r1, r2, r3$ are all not i. If $V_i^{Gen} \notin [X_{\min}, X_{\max}]$, then regenerate a mutation individual V_i^{Gen} from Formula (20).

(3) Recombination: before and after mutation, each property element of an individual undergoes recombination and produces a new individual U_i^{Gen}. The rule of recombination is as follows:

$$U_i^{Gen}(x_j) = \begin{cases} V_i^{Gen}(x_j), \, rand(0,1) < CR \, or \, randn = j \\ X_i^{Gen}(x_j), \, others \end{cases}, j = 1, 2 \qquad (22)$$

where CR is a recombination constant, generally $CR \in [0.8,1]$, $U_i^{Gen}(x_j)$ refers to the property value of x_j in U_i^{Gen}, and $randn$ is a random value of 1 or 2.

(4) Selection: superior individuals replace an original individual and go into the next generation. Select superior individuals according to Formula (23):

$$X_i^{Gen+1} = \begin{cases} U_i^{Gen}, \, F(U_i^{Gen}) < F(X_i^{Gen}) \\ X_i^{Gen}, \, F(U_i^{Gen}) > \, = F(X_i^{Gen}) \end{cases} \qquad (23)$$

According to the given steps, perform cycle evolution and iterative computation until adaptability meets requirements (the value of target function meets limited conditions, for example, it can be controlled by $r \geq 0.95$) or ends when it evolves to a maximum algebra maxGen. After that, select the individual having the minimum value of fitness function (the parameter values that individual elements stand for are x_0 and t_0 to be determined), and the calculation is completed.

4.3 Improvement to coupled probability density method

The model uses the DEA for solution, the result of calculation has a great randomness, and some optimized results might appear repeatedly. Therefore, some optimized rules should be added to control the value of optimized fitness function to make it change unidirectionally. For this, this chapter puts forward the application of gradient concept in a differential evolution process to provide an optimization direction and improving optimization efficiency.

Firstly, add two new attributes signifying gradient into an individual, namely, gradient feature factor $gc(X)$ and gradient direction factor $gp(X)$. A population individual newly defined is $S_i^{Gen} = S_i^{Gen}(X_i^{Gen}, gc(X_i^{Gen}), gp(X_i^{Gen}))$, where the gradient feature factor $gc(X_i)$ and gradient direction factor $gp(X_i)$ are determined from Formulas (24) and (25).

$$gc(X_i^{Gen}) = \begin{cases} 1, & F(X_i^{Gen}) \leq F(X_i^{Gen-1}) \\ 0, & F(X_i^{Gen}) > F(X_i^{Gen-1}) \end{cases} \qquad (24)$$

$$gp(X_i^{Gen}) = \begin{cases} [sgn(X_i^{Gen}(x_1) - X_i^{Gen-1}(x_1)), sgn(X_i^{Gen}(x_2) - X_i^{Gen-1}(x_2))], \, gc(X_i^{Gen}) = 1 \\ 0, \, gc(X_i^{Gen}) = 0 \end{cases}$$
$$(25)$$

where sgn is a symbolic function.

After setting gradient attributes, the trend information that facilitates population evolution can be kept during differential evolution selection. For example, supposing an individual X_i^{Gen} is superior to its previous generation of individuals X_i^{Gen-1} and $gp(X_i^{Gen}) = [1, -1]$, then this means that the transient superior evolution trend of this individual is "increasing position parameter x_0 and reducing time

parameter t_0." The information on such trend can be used to guide the evolution of the next generation of individuals. When making improvement to DEA optimization in actual applications, we can utilize and obtain gradient information as an additional rule to control evolution direction and guide the evolution of each generation of individuals.

5. Application examples

As shown in **Figure 3**, a emergent treatment demonstration project of sudden water pollution events was carried out from the check gate in the Fangshui River at Jingshi Section (Pile 100 + 294.750) to the check gate in the Puyang River (Pile No. 113 + 492.750) of the SNWDMRP on March 22, 2014. Cane sugar was used as tracer, and the concentration of cane sugar solution, as a water quality detection index. The experimental channel has a trapezium cross section, whose parameters are bottom width = 18.5 m, side slope = 2.5, average water depth = 4.0 m, channel bottom slope = 0.00005, manning roughness = 0.015 (design value), the experimental estimated value of dispersion coefficient = 3.43 m²/s, and constant flow ≈ 8.0m³/s. About 800 kg of cane sugar was put instantaneously into the channel at 9:00 AM on March 22, and four monitoring cross sections were set downstream. In order to ensure successful instantaneous putting of 800 kg cane sugar, the cane sugar was dissolved in heated clear water of 1.0 m³ (about 80.0°C) timely and in advance. When the experiment started, a floating bridge already built was used to directly pour dissolved cane sugar solution into the middle of the channel. The concentration monitoring process at 1508 m downstream of the cross section where the solution was poured is shown in **Figure 4**, and the start time of monitoring was 9:30.

Supposing the pile number of the pouring cross section is 0 m, then the pile number of the monitored cross section is 1508 m. The width and water depth at the experimental section of the channel were very small with regard to the length. After checking calculation, the pollutants 1508 m from the monitored cross section were uniformly mixed roughly in the lateral and vertical directions. The time for urgently coping with sudden pollution is short. Without considering degradation, the transport of pollutants can be described by using a convective diffusion equation at a 1D constant flow. A traceability model is applied for traceability calculation,
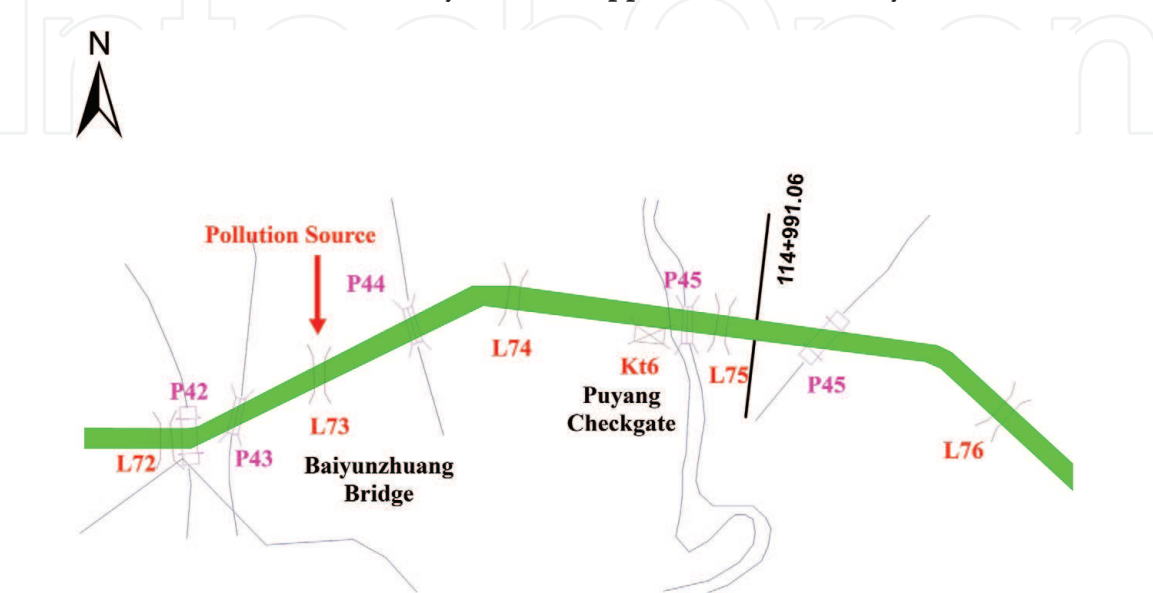

Figure 3.
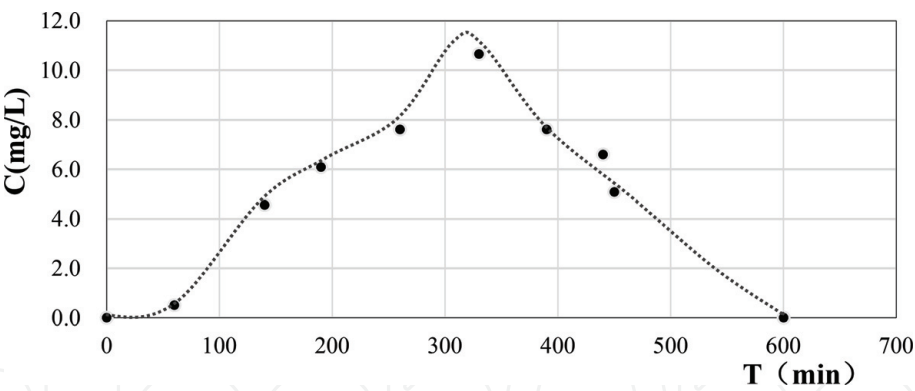*Schematic diagram of experiment channel selected for demonstration project.*

Figure 4.
Actual concentration monitoring process at monitored cross section.

	M (kg)	x_0 (m)	t_0
Actual value	800.00	0.00	9:00
The value calculated by the probabilistic method	781	−570	9:14 AM
The value calculated by the coupled density method	731.13	−28.42	8:48 AM

Table 1.
Table of results of traceability by all methods.

	M (kg)	x_0 (m)	t_0
The error calculated by the probabilistic method	2.38%	570	14 min
The error calculated by the coupled probability density method	8.59%	28.42	12 min

Table 2.
Analysis of errors of traceability by all methods.

which was established by using the probability statistics method and the proposed coupled probability density method among conventional traceability methods. The results are shown in **Tables 1** and **2**.

From **Tables 1** and **2**, we can see that the error in the position of a pollution source calculated by using the coupled probability density method is not up to 30 m, time error about 12 min, and the intensity of pollutants not larger than 10%. However, considering the calculation done by the probability statistics method, although the precision of its pollutant intensity is higher, the position and time are not calculated more accurately, especially the error of the position calculated is above 500 m. Furthermore, to solve the same problem, it takes much more time for calculation by the probability statistics method than by the probability density method. As a whole, when a sudden water pollution accident happens during water diversion, the adoption of the traceability technology based on the coupled probability density method is recommended more.

6. Summary

Due to uncertainty of sudden water pollution events, after occurrence of an event, it is very difficult to determine the location and release time by experience. Therefore, the first task for emergent treatment is to determine the pollution source. The main content of fast traceability of sudden pollution is to determine the

location, release time, and the intensity of the pollution source. Regarding fast determination of the conditions of a pollution source in case of sudden water pollution under SNWDMRP, this chapter presents detailed introduction and analysis of the fast channel traceability technology including conventional traceability, provides practical effect of the traceability technology via analysis of example applications, and recommends the application of the coupled probability density method to traceability calculation according to traceability results, in order to provide support for fast traceability and emergent regulation and control of sudden water pollution under long-distance water transfer projects.

However, the traceability model can have good application results for single-point source of sudden water pollution accident in a river. The traceability model for multipoint sources of sudden water pollution accidents needs deeper research and can refer to the ongoing research by Wang [19].

Author details

Weihong Liao[1], Zhiguo Gan[1], Jiabiao Wang[2]* and Hezhen Zheng[3,4]

1 China Institute of Water Resources and Hydropower Research, State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, Beijing, China

2 Tsinghua University, Beijing, China

3 Zhejiang University, Hangzhou, China

4 Changjiang Survey, Planning, Design and Research Co., Ltd. Wuhan, China

*Address all correspondence to: waterwhu@foxmail.com

IntechOpen

References

[1] Wang J, Lei X, Liao W, Wang H. Source identification for river sudden water contamination based on coupled probability density function method. Shui Li Xue Bao. 2015;**46**(11): 1280-1289

[2] Cheng WP, Liao XJ. Recovering the release history of contaminant based on backward probability density function for the one-dimensional convection diffusion equation. Chinese Journal of Hydrodynamics. 2011;**26**(4):460-469

[3] Ghane A, Mazaheri M, Mohammad VSJ. Location and release time identification of pollution point source in river networks based on the backward probability method. Journal of Environmental Management. 2016; **180**:164-171

[4] Yang H, Shao D, Liu B, Huang J, Ye X. Multi-point source identification of sudden water pollution accidents in surface waters based on differential evolution and Metropolis–Hastings–Markov chain Monte Carlo. Stochastic Environmental Research and Risk Assessment. 2016;**30**(2):507-522

[5] Xu T, Gomez-Hernandez JJ. Joint identification of contaminant source location, initial release time and initial solute concentration in an aquifer via ensemble Kalman filtering. Water Resources Research. 2016;**52**(8): 6587-6595

[6] Skaggs TH, Kabala ZJ. Recovering the release history of a groundwater contaminant. Water Resources Research. 1994;**30**(1):71-79

[7] Akçelik V, Biros G, Ghattas O, Long KR, Waanders BB. A variational finite element method for source inversion for convective-diffusive transport. Finite Elements in Analysis and Design. 2003; **39**(8):683-705

[8] Han L, Jiang L, Zhu D. Inverse problem on boundary condition and pollutant source in water quality model of river network. Journal of Hohai University. 2001;**29**(5):23-26

[9] Gorelick SM, Evans B, Remson I. Identifying sources of groundwater pollution: An optimization approach. Water Resources Research. 1983;**19**(3): 779-790

[10] Alapati S, Kabala ZJ. Recovering the release history of a groundwater contaminant using a non-linear least-squares method. Hydrological Processes. 2000;**14**(6):1003-1016

[11] Cao X, Song J, Zhang W, Zhang L. MCMC method on an inverse problem of source term identification for convection-diffusion equation. Chinese Journal of Hydrodynamics. 2010;**25**(2): 127-136

[12] Hazart A, Giovannelli JF, Dubost S, Chatellier L. Inverse transport problem of estimating point-like source using a Bayesian parametric method with MCMC. Signal Processing. 2014;**96**(5): 346-361

[13] Jin ZQ, Chen XQ. Numerical solution to an inversed problem of source control for convection-diffusion equations by PST-optimization. Journal of Hohai University. 1992;**20**(2):1-8

[14] Chen YH, Wang P, Jiang JP, Guo L. Contaminant point source identification of rivers chemical spills based on correlation coefficients optimization method. China Environmental Science. 2011;**31**(11):1802-1807

[15] Min T, Zhou XD, Zhang SM, Feng MQ. Genetic algorithm to an inverse problem of source term identification for convection-diffusion equation.

Journal of Hydrodynamics, Ser. A. 2004;**19**(4):520-524

[16] Neupauer RM, Wilson JL. Adjoint method for obtaining backward-in-time location and travel time probabilities of a conservative groundwater contaminant. Water Resources Research. 1999;**35**(11):3389-3398

[17] Cheng W, Jia Y. Identification of contaminant point source in surface waters based on backward location probability density function method. Advances in Water Resources. 2010;**33**(4):397-410

[18] Storn R, Price K. Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization. 1997;**11**(4):341-359

[19] Wang JB, Zhao JS, Lei XH, Wang H. New approach for point pollution source identification in rivers based on the backward probability method. Environmental Pollution. 2018;**241**: 759-774